A lattice Dirac operator
for QCD with light dynamical quarks

Nigel Cundy[a,c], A D Kennedy[b], Andreas Schäfer[a]

[a]*Institut für Theoretische Physik, Universität Regensburg, D-93040 Regensburg, Germany*
[b]*SUPA, School of Physics & Astronomy,*
University of Edinburgh, Edinburgh EH9 3JZ, Scotland
[c]*Lattice Gauge Theory Research Center, FPRD, and CTP,*
Department of Physics & Astronomy, Seoul National University, Seoul, 151-747,
South Korea

Abstract

In QCD chiral symmetry is explicitly broken by quark masses, the effect of which can be described reliably by chiral perturbation theory. Effects of explicit chiral symmetry breaking by the lattice regularisation of the Dirac operator, typically parametrised by the residual mass, should be negligible for almost all observables if the residual mass of the Dirac operator is much smaller than the quark mass. However, maintaining a small residual mass becomes increasingly expensive as the quark mass decreases towards the physical value and the continuum limit is approached. We investigate the feasibility of using a new approximately chiral Dirac operator with a small residual mass as an alternative to overlap and domain wall fermions for lattice simulations. Our Dirac operator is constructed from a Zolotarev rational approximation for the matrix sign function that is optimal for bulk modes of the Hermitian kernel Dirac operator but not for the low-lying parts of its spectrum. We test our operator on various $32^3 \times 64$ lattices, comparing the residual mass and the performance of the Hybrid Monte Carlo algorithm at a similar lattice spacing and pion mass with a hyperbolic tangent operator as used by domain wall fermions. We find that our approximations have a significantly smaller residual mass than domain wall fermions at a similar computational cost, and still admit topological charge change.

1. Introduction

Lattice QCD is a phenomenologically successful regularisation of QCD which can accurately predict experimental observables. It places space-time on a discrete lattice, and the continuum theory is recovered as three limits are taken, the volume to infinity, the lattice spacing to zero, and the quark masses to their physical values. The extrapolations in the lattice spacing and quark mass are made more difficult, with increased errors, by explicit breaking of chiral

symmetry within the lattice discretization of the Dirac operator. The Nielson-Ninomiya theorem [1], which states that it is impossible to represent an odd number of quark flavours on the lattice without breaking chiral symmetry is avoided for the cheapest commonly used families of lattice actions either by simulating additional degenerate fermions (staggered fermions) or by explicitly breaking chiral symmetry (Wilson fermions).

There is, however, an alternative, to use a lattice chiral symmetry [2], defined by a Ginsparg-Wilson relation [3], which has a smooth limit to the continuum chiral symmetry. The only known and practical lattice Dirac operators which satisfy lattice chiral symmetry are discontinuous, being built from the matrix sign function. The simplest of these is the overlap operator, first proposed by Neuberger and Narayanan [4, 5]. However, the computational difficulties involved in simulating overlap fermions mean that their use in lattice simulations remains challenging. A previous solution with a a good approximate chiral symmetry, the domain wall fermion, which placed left handed and right handed fermions on the four dimensional surfaces at the boundaries of a five dimensional lattice [6]. This formalism reduces to a form of the overlap operator for infinite fifth dimension, giving exactly chiral fermions. It is also equivalent to the four dimensional Kenney-Laub-Neuberger [7, 5] Dirac operator.

There are, however, many other ways in which an approximate chiral symmetry can be maintained on the lattice. There are no *a priori* reasons to suspect that the historical domain wall fermion should be the best (the 'best' implementation can be defined as the one that requires the least computational effort for a particular residual mass, the standard measurement of explicit chiral symmetry breaking on the lattice). In this work, we propose and investigate a four dimensional approximation to the overlap operator that gives a considerably better residual mass than domain wall fermions while avoiding the complexities of exact overlap fermions.

1.1. On-Shell Chiral Symmetry

For a lattice Dirac operator $\tilde{\slashed{D}}$, an on-shell chiral transformation can be written as [2]

$$\psi \mapsto e^{i\alpha\gamma_5(1-a\tilde{\slashed{D}})}\psi \qquad \text{and} \qquad \bar{\psi} \mapsto \bar{\psi}e^{i\alpha(1-a\tilde{\slashed{D}})\gamma_5};$$

the differential condition that the Dirac operator itself is invariant under this transformation is a Ginsparg–Wilson relation [3]

$$\{\gamma_5, \tilde{\slashed{D}}\} = 2a\tilde{\slashed{D}}\gamma_5\tilde{\slashed{D}},$$

which may be written in the equivalent forms

$$\text{Re}\,\tilde{\slashed{D}} = a\tilde{\slashed{D}}^\dagger\tilde{\slashed{D}} \quad \Leftrightarrow \quad \text{Re}[(1-a\tilde{\slashed{D}})^\dagger a\tilde{\slashed{D}}] = 0 \quad \Leftrightarrow \quad \text{Re}\,\tilde{\slashed{D}}^{-1} = a.$$

We may define the quantity $\hat{\gamma}_5$ by $a\tilde{\slashed{D}} = \frac{1}{2}(1+\gamma_5\hat{\gamma}_5)$, and if we wish $\tilde{\slashed{D}}$ to satisfy $\tilde{\slashed{D}}^\dagger = \gamma_5\tilde{\slashed{D}}\gamma_5$ then $\hat{\gamma}_5 = \hat{\gamma}_5^\dagger$. This Ginsparg–Wilson relation implies that $\hat{\gamma}_5^2 = 1$,

so $\hat{\gamma}_5$ is both hermitian and unitary, so that its spectrum can contain only the eigenvalues ± 1 and hence $\hat{\gamma}_5 = \mathrm{sgn}(\hat{\gamma}_5)$.

We must also require that $\tilde{\slashed{D}}$ is a Dirac operator in the naïve continuum limit, $\tilde{\slashed{D}} = Z_{GW}(\slashed{\partial} + \slashed{A}) + O(a^2)$, where there are no corrections of $O(a)$ because the only available such operator is $[\slashed{\partial} + \slashed{A}, \slashed{\partial} + \slashed{A}] = \sigma \cdot F \equiv \sigma_{\alpha\beta} F^{\alpha\beta}$ which is not chirally invariant. If $\slashed{D}_W$ is any γ_5-hermitian lattice Dirac operator satisfying $\slashed{D}_W = Z_W(\slashed{\partial} + \slashed{A}) + O(a)$ then we may choose $2aM\tilde{\slashed{D}} = \slashed{D}_W + O(a)$ where $2aM = Z_W/Z_{GW}$ is a suitable finite wave-function renormalization, and we obtain the (massless) Neuberger operator [8, 9]

$$a\tilde{\slashed{D}}_0 \equiv \tfrac{1}{2}(1 + \gamma_5 \, \mathrm{sgn} \, H) \tag{1}$$

where $H \equiv \gamma_5 \slashed{D}_W - M$. In QCD a quark has a (small) mass μ, and this can be incorporated by shifting the spectrum to lie in the interval $[a\mu, 1]$ by defining the massive Neuberger operator to be

$$a\tilde{\slashed{D}}\mu \equiv \tfrac{1}{2}[(1 + a\mu) + (1 - a\mu)\gamma_5 \, \mathrm{sgn}(H)]. \tag{2}$$

Within this framework there are numerous choices that have to be made [10]: whether to use four or five dimensional pseudofermions; the choice of the kernel operator H (including the mass M); the choice of rational approximation for the matrix sign function; which form of 5D matrix with the desired Schur complement (for example, Euclidean-Caley or continued fraction); and for four dimensional pseudofermions whether to use a nested 4D or 5D inverter. All of these choices are independent and physically equivalent. We use the name "domain wall operator" as any approach which uses five dimensional pseudofermions, and "overlap operator" as any approach in four dimensions with an exact matrix sign function, limited only by the floating point precision of the computer. An "approximate overlap operator" is a four dimensional approach which has a small explicit breaking of chiral symmetry from the use of an approximate matrix sign function. In this article, we are only studying the effect of changing the rational approximation used for the matrix sign function. Our results should be independent of the choice of kernel and whether four dimensional or five dimensional pseudofermions are used.

1.2. Ginsparg–Wilson Defect and Residual Mass

Suppose we have some approximation $\varepsilon(H) \approx \mathrm{sgn}(H)$ to the matrix sign function, so that equation (2) is replaced by

$$D_\mu = \frac{1}{2}[(1 + a\mu) + (1 - a\mu)\varepsilon(H)], \tag{3}$$

then we may define the defect Δ as the amount by which the corresponding approximate Neuberger operator, $\slashed{D}$, fails to satisfy the Ginsparg–Wilson relation for $\mu = 0$,

$$\Delta \equiv \tfrac{1}{2}\{\gamma_5, \slashed{D}_0\} - a\slashed{D}_0\gamma_5\slashed{D}_0 = \gamma_5 \, \mathrm{Re}[(1 - a\slashed{D}_0)^\dagger\slashed{D}_0],$$

3

which gives

$$4\gamma_5 a\Delta = 1 - \varepsilon(H)^2. \qquad (4)$$

From equation (3) we have $\slashed{D}_\mu = \slashed{D}_0 + \mu(1 - a\slashed{D}_0)$, hence

$$
\begin{aligned}
\mathrm{Re}[(1 - a\slashed{D}_0)^\dagger \slashed{D}_\mu] &= \mathrm{Re}[(1 - a\slashed{D}_0)^\dagger \slashed{D}_0] + \mu(1 - a\slashed{D}_0)^\dagger(1 - a\slashed{D}_0) \\
&= \gamma_5\Delta + \mu(1 - a\slashed{D}_0)^\dagger(1 - a\slashed{D}_0).
\end{aligned}
$$

Multiplying this by $\slashed{D}_\mu^{\dagger\,-1}$ on the left and $\slashed{D}_\mu^{-1}$ on the right we obtain

$$\tfrac{1}{2}(S_\mu + S_\mu^\dagger) = \slashed{D}_\mu^{\dagger\,-1}\gamma_5\Delta\slashed{D}_\mu^{-1} + \mu S_\mu^\dagger S_\mu; \qquad (5)$$

where $S_\mu \equiv (1 - a\slashed{D}_0)\slashed{D}_\mu^{-1}$ (which satisfies $S_\mu^{-1} = S_0^{-1} + \mu$). The first term on the right side of equation (5) is a measure of the chiral symmetry breaking due to the approximation to the sign function whereas the second is that due to the explicit quark mass $a\mu$. We wish to introduce some norm $\|\Delta\|$ on the defect to quantify the magnitude of the errors due to our approximation to the sign function. A useful estimate for this norm is the residual mass

$$m'_{\mathrm{res}} \equiv \frac{\mathrm{tr}(\slashed{D}_\mu^{\dagger\,-1}\gamma_5\Delta\slashed{D}_\mu^{-1})}{\mathrm{tr}(S_\mu^\dagger S_\mu)},$$

whence $\mathrm{tr}(S_\mu + S_\mu^\dagger)/2\,\mathrm{tr}(S_\mu S_\mu^\dagger) = m_{\mathrm{res}} + \mu$.

There is no a priori reason when comparing the extent of chiral symmetry breaking for different Dirac operators why we should project the defect onto a scalar using this trace, indeed in our numerical studies we have used the computationally cheaper approach of projecting onto momentum zero states, giving

$$m_{\mathrm{res}} = \frac{\sum_{x,x'}\left[\slashed{D}_\mu^{\dagger\,-1}\gamma_5\Delta\slashed{D}_\mu^{-1}\right]_{x,x'}}{\sum_{x,x'}\left[S_\mu^\dagger S_\mu\right]_{x,x'}}.$$

We do not expect that the quantity m_{res} will be less suitable than m'_{res} to compare the effects of chiral symmetry breaking between different Dirac operators.

Since our approximation for the Neuberger operator does not exactly satisfy chiral symmetry it will have $O(a)$ corrections near the continuum limit, and as these must either come from the explicit mass term or the defect we see that

$$\slashed{D}_\mu = Z_{GW}(\slashed{\partial} + \slashed{A}) + \mu + m_{\mathrm{res}} + \text{constant} \times a\sigma \cdot F + O(a^2). \qquad (6)$$

1.3. Effects of Residual Mass

The spectrum of the exact unitary matrix $\gamma_5\hat{\gamma}_5 = \gamma_5\,\mathrm{sgn}(H)$ lies on the unit circle in the complex plane $\|\gamma_5\hat{\gamma}_5\| = 1$. The approximate matrix sign function has spectral norm $\|\varepsilon(H)\| \equiv \sup_{|\psi|=1}\psi^\dagger\varepsilon(H)\psi$, so the spectrum of $\|\gamma_5\varepsilon(H)\| \leq \|\gamma_5\|\,\|\varepsilon(H)\| = \|\varepsilon(H)\|$ lies within a disc of radius $\|\varepsilon(H)\|$. For an operator with good chiral symmetry most eigenvalues will be close to the

unit circle; however small eigenvalues of H, where the approximation is less good, may exhibit large discrepancies from the spectrum of the exact overlap operator. For example, if H has an exactly zero eigenvalue then for all symmetric approximations to the matrix sign function the corresponding eigenvalue of $\not{D}_\mu$ will lie in the centre of the circle, i.e. at $\frac{1}{2}(1 + a\mu)$.

If the residual mass is sufficiently smaller than the target physical quark mass, $m_{\mathrm{res}} \ll \mu$ then as long as the lattice spacing is small enough that the $O(a^2)$ effects in equation (6) are negligible all physical effects of m_{res} can be removed by adjusting $\mu \mapsto \mu - m_{\mathrm{res}}$ so that the quark mass stays fixed. However, if the residual mass is larger than the physical quark mass then this is not possible in general. While one could insert $\mu < 0$ into the lattice Dirac operator this would be likely to introduce zero or negative eigenvalues, leading to the inexact overlap Dirac operator becoming singular on some (now exceptional) configurations.

It is therefore desirable to have a small residual mass. The difficulty with this is that equation (4) tells us that $\|\Delta\| = \|1 - \varepsilon(H)\|/4a$, so we need to reduce the error $\|1 - \varepsilon(H)\| \propto a$ in our approximation to keep the physical m_{res} fixed as the continuum limit is approached. This can be done by increasing the order of the rational approximation which may, depending on the approximation used, significantly increase the cost of the simulation.

At larger lattice spacings it is easier to maintain a small m_{res}, but there will be larger $O(a^2)$ and higher lattice artefacts in equation (6); since these come from the chiral symmetry breaking in Δ and not just from the quark mass μ they are less easy to model using, for example, chiral perturbation theory.

1.4. Choice of Rational Approximation

We have shown that we need a lattice Dirac operator with a good approximation to (on-shell) chiral symmetry, but we also wish to avoid the cost of maintaining exact chiral symmetry to machine precision. Our goal is thus to find a family of approximations to the matrix sign function that provides a good balance between residual mass and the cost of Hybrid Monte Carlo (HMC) computations. We shall demonstrate that domain wall fermions, where the accuracy of whose hyperbolic tangent approximation falls too slowly with increasing rational degree are far from optimal, while overlap fermions, which are far more chiral than necessary for massive quarks at a cost in time and complexity of the HMC algorithm, are more expensive than is required to obtain a good enough chiral symmetry.

In this work, we introduce and test the Zolotarev lattice Dirac operator, which uses the optimal rational approximation to the matrix sign function, but not over the entire spectrum of the kernel operator $\not{D}_W$. This is guaranteed to provide the best approximation to the matrix sign function for a given order of rational approximation within a certain tunable eigenvalue range, and therefore might be expected to give the smallest residual mass for a given amount of computational effort. However, there are a number of questions which need to be addressed in this comparison.

1. While the Zolotarev approximation is guaranteed to give the best L_∞ approximation over part of the spectrum, some eigenvalues lie outside this range. What contribution do they make to the residual mass?
2. Does the reduced residual mass come at the cost of less stable molecular dynamics? If a small time step were required to resolve larger fluctuations in the fermionic force this could compensate for any gain in the residual mass.
3. Is the Zolotarev Dirac operator ergodic? In particular, are all topological sectors sampled, and is the autocorrelation between topological sectors and different portions of the same topological sectors short enough?
4. Is the Zolotarev Dirac operator local? As it does not approximate the sign function well for small eigenvalues of the kernel operator it is not obvious that we can rely on previous proofs and numerical results, although it seems unlikely that it could be less local than the hyperbolic tangent Dirac operator used in domain wall computations.

In §2 we describe our Zolotarev approximation and compare it to the Kenney–Laub–Neuberger [7, 5] hyperbolic tangent approximation, which is a four dimensional representation of the five dimensional domain wall Dirac operator; in §3 we describe how we tested the various methods; in §4 we present the numerical results from these tests; and in §5 we present our conclusions.

2. The Zolotarev Dirac operator

2.1. The Zolotarev approximation

The Zolotarev Dirac operator depends on the degree N of the rational approximation and a parameter ξ where $\xi\|H\| \leq |\lambda| \leq \|H\|$ is the interval of the kernel operator's spectrum on which the approximation to the matrix sign function is optimal. Here $\|H\| = \sup_{\|u\|_2=1} (u, Hu)$ is the spectral norm of the hermitian kernel operator H, i.e., its largest eigenvalue. The Zolotarev Dirac operator is

$$a\slashed{D}_Z = \tfrac{1}{2}\left[(1 + a\mu) + (1 - a\mu)\gamma_5\, \varepsilon_{\mathrm{z}}\left(\frac{H}{\|H\|}\right)\right]$$

where

$$\varepsilon_{\mathrm{z}}(x) = x \sum_{i-1}^{\lfloor N/2\rfloor} \frac{\omega_i}{x^2 - \sigma_i^2}$$

is the Zolotarev approximation to the sign function, which is optimal in the L_∞ norm over $\xi \leq |x| \leq 1$. The coefficients ω_i and σ_i can be computed in terms of Jacobi elliptic functions[1] depending upon ξ and N [11, 12, 13, 14]. Were the

---

[1] In particular $\sigma_j = \xi\,\mathrm{sn}\left(2i\,\mathrm{K}'(j - \tfrac{1}{2})/N, \xi\right)$, where $\mathrm{K}'(k) = \mathrm{K}\left(\sqrt{1 - k^2}\right)$ is a complete elliptic integral.
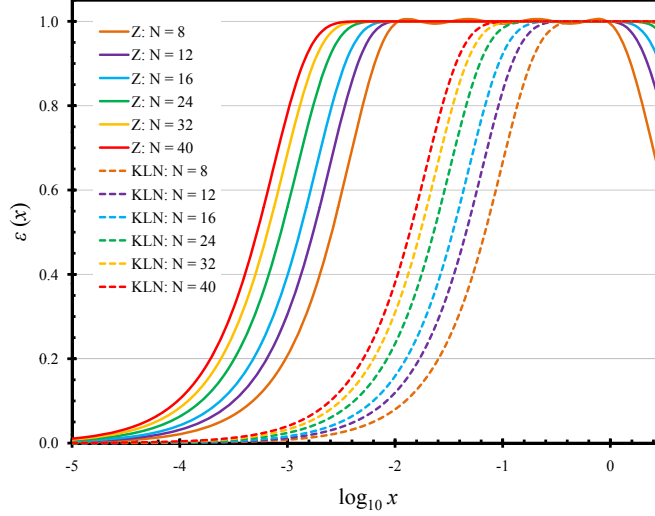
Figure 1: The solid lines show Zolotarev rational approximations $\varepsilon_Z(x)$ for various degrees N with $\xi = 10^{-2}$ as a function of $\log_{10} x$. For comparison the dashed lines show the Kenney–Laub–Neuberger hyperbolic tangent approximations $\varepsilon_{\mathrm{KLN}}(x)$ of the same degrees.

Zolotarev rational approximation expressed in a five dimensional formalism, similar to that used for domain wall fermions, the size of the fifth dimension would be $L_s = N$.

There are two families of Zolotarev approximations, those that vanish at the origin and those that are singular at the origin (up to a small re-scaling the latter are just the reciprocals of the former). We shall only use the non-singular kind, which vary smoothly and monotonically from -1 just below $-\xi$ to $+1$ just above ξ. Figure 1 shows the Zolotarev approximation $\varepsilon_Z(x)$ for various degrees N and for a typical value $\xi = 0.01$. The maximum error of ε_Z over the interval on which it is optimal, $\Delta = \max_{\xi \leq |x| \leq 1} |\varepsilon_Z(x) - 1|$, is shown in Figure 2. Not only can we see from Figure 2 that the error in the approximation Δ falls exponentially with the degree N over the interval where the Zolotarev approximation is optimal, but also from Figure 1 we see that the approximation improves over the interval $|x| < \xi$.

2.2. The KLN Dirac operator

The Kenney–Laub–Neuberger Dirac operator operator (KLN), which is implicitly used in the domain wall approach [10, 15, 16], is

$$a\slashed{D}_{\mathrm{KLN}} = \tfrac{1}{2}\left[(1 + a\mu) + (1 - a\mu)\gamma_5\,\varepsilon_{\mathrm{KLN}}\left(\frac{H}{\|H\|}\right)\right]$$

7

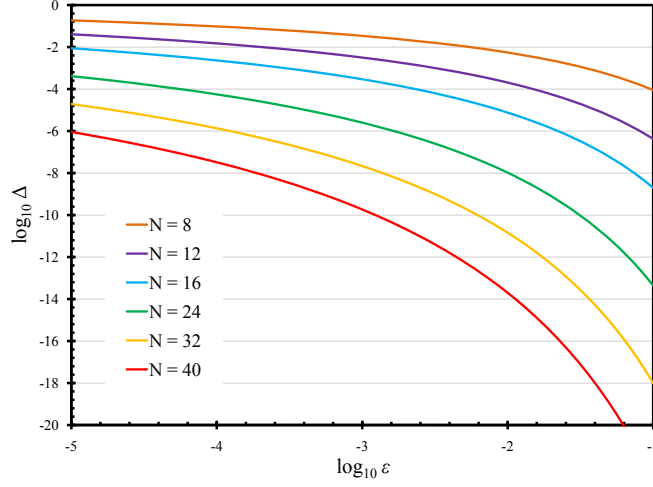Figure 2: The maximum error of the Zolotarev approximation $Z(x)$ over its approximation interval, $\Delta = \max_{\xi \le |x| \le 1} |\varepsilon_Z(x) - 1|$ for various degrees N are shown on a log-log plot.

where the Kenney–Laub–Neuberger [7, 5] hyperbolic tangent approximation to the sign function is

$$\varepsilon_{\text{KLN}}(x) = \tanh\left(N \tanh^{-1} x\right) = \frac{(1+x)^N - (1-x)^N}{(1+x)^N + (1-x)^N}.$$

This may be written for even N as a partial fraction expansion [10]

$$\varepsilon_{\text{KLN}}(x) = \frac{2x}{N} \sum_{k=0}^{\frac{N}{2}-1} \frac{1 + \left(\tan \frac{(k+\frac{1}{2})\pi}{N}\right)^2}{x^2 + \left(\tan \frac{(k+\frac{1}{2})\pi}{N}\right)^2}.$$

Figures 1 shows how this approximation compares to the Zolotarev approximation. Unlike the Zolotarev approximation it does not require a minimum eigenvalue as an input; like the Zolotarev approximation the accuracy depends on the order of the rational approximation.

2.3. Fermionic forces in the Hybrid Monte Carlo algorithm

One of the major possible difficulties with this method is that a very small integration step size $\delta\tau$ may be required to prevent the molecular dynamics (MD) trajectory in the Hybrid Monte Carlo (HMC) [17] algorithm from becoming unstable. For overlap fermions these instabilities can be avoided (albeit with a little additional complexity in the force calculation [18]) because the small eigenvalues are not treated by an approximation but deflated, and the simulation is run in the chiral sector without zero modes [19, 20]. Here, because

8

all the eigenvalues are treated by the rational approximation, the force acting on the gauge fields from their interaction with fermion fields might become large for three reasons:

1. the fermion mass is small and the gauge-pseudofermion coupling involves the inverse of the Dirac operator which will, in general, have approximate zero modes;
2. the derivative of the chiral Dirac operator $\slashed{D}_\mu$ becomes large if the fermion field is close to a zero mode of the kernel operator H (for an exact overlap operator, it is a Dirac δ-function); and
3. the estimate of the fermionic force obtained from pseudofermion fields is noisy [21, 22].

The use of multiple pseudofermion fields [23, 24] can reduce the effects of (3) and, for overlap fermions, a factorisation of the determinant can can completely remove the effect of the pseudofermion noise [22]. In overlap simulations, various transmission–reflection methods [25, 26, 20] have been introduced to resolve the Dirac δ-function from the effects of point 2 above. However, the adaptations to the usual HMC algorithm needed for exact overlap fermions are expensive, so for the case of interest here — where chiral symmetry is broken explicitly by a small fermion mass $a\mu$ — we advocate choosing a Zolotarev approximation that is not optimal for the smallest eigenvalues of H, but is a compromise between having a small residual mass $m_{\mathrm{res}} \ll \mu$ and having a small fermionic force. It is clear from Figure 1 that we may expect the Zolotarev Dirac operator to have a much smaller residual mass than the KLN Dirac operator used in the domain wall method for any degree N and reasonable values of ξ.

The fermionic force for continuous time MD evolution, to which the discrete integrators used in HMC are a good approximation for reasonable acceptance rates, contains a term proportional to the derivative of the approximation to the sign function used in $\slashed{D}_\mu$ [18]. The derivatives of the Zolotarev and KLN approximations are shown in Figure 3, and several interesting features are immediately obvious. For both Zolotarev and KLN approximations the largest derivative occurs at $x = 0$, this is trivial for the KLN approximation where the derivative at the origin is $\varepsilon'_{\mathrm{KLN}}(0) = N$, but it is less obvious for the Zolotarev approximation. We first note that the derivative $\varepsilon'_{\mathrm{z}}(x)$ over the interval $\xi \leq |x| \leq 1$ on which the approximation is optimal is always small, and falls exponentially with the degree N. Next we see that the derivative increases monotonically as x approaches zero where it attains its maximum value. In Figure 4 we show the dependence of $\varepsilon'_{\mathrm{z}}(0)$ on ξ and the degree N; empirically the data are well fitted by $\varepsilon'_{\mathrm{z}}(0) \approx (0.49 + 0.025N)/\xi^{0.87}$.

3. Numerical Implementation

3.1. Markov Chains

Our goal in this work is to investigate whether the Zolotarev Dirac operator gives a smaller residual mass for the same cost as the domain wall Dirac operator. To do this, we have generated several small ensembles on $32^3 \times 64$ lattices
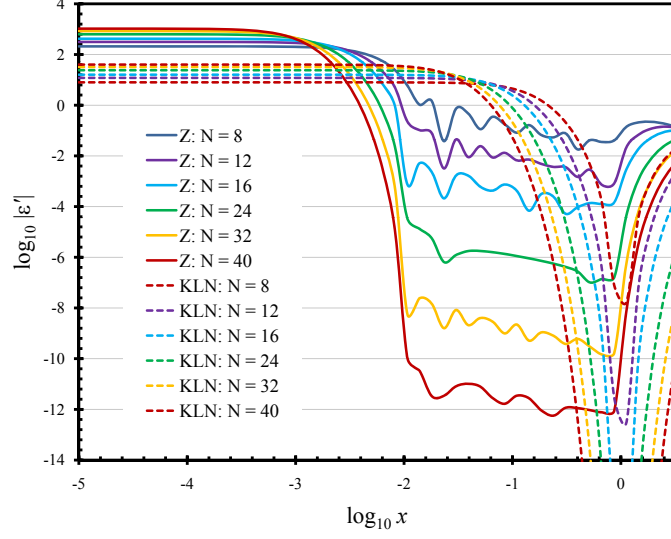
Figure 3: The derivative of the Zolotarev rational approximation $|d\varepsilon_{\mathrm{Z}}(x)/dx|$ with $\xi = 10^{-2}$ and of the Kenney–Laub–Neuberger hyperbolic tangent approximation $|d\varepsilon_{\mathrm{KLN}}(x)/dx|$ for various orders N.
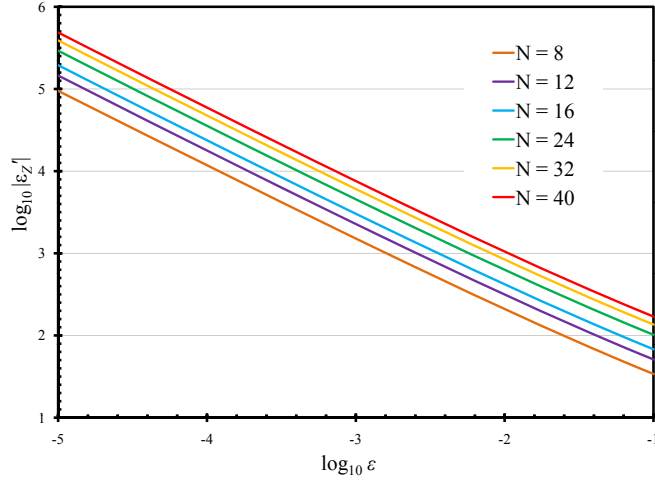

Figure 4: The derivative of the Zolotarev rational approximation $|d\varepsilon_{\mathrm{Z}}/dx|$ at the origin $x = 0$ for various orders N.

starting from equilibrated domain wall configurations. We aimed to match the pion masses and lattice spacings for these ensembles, although in practice the Zolotarev ensembles were at a slightly lighter pion mass and slightly smaller lattice spacing. We compared the KLN Dirac operator, which is equivalent to a domain wall fermion with a Boriçi kernel, with three different Zolotarev Dirac operators with different values of ξ. We used a Boriçi–Wilson kernel with one step of over-improved [27] stout smearing [28], at $\kappa = 0.19$ and a tadpole improved Lüscher-Weisz gauge action [29, 30, 31, 32]. We ran enough trajectories in each case to ensure that the plaquette was thermalised with the new action before taking measurements of the pion mass, lattice spacing, and residual mass. There is a small systematic uncertainty in these measurements as our ensembles were not fully thermalized with respect to the measured observables, however we have noticed no change in our measurements along the Markov chains, so any systematic error due to incomplete thermalization is probably smaller than our statistical errors. We have made measurements on 5–10 configurations for each run. Our goal in making these measurements is not to extract physics, but to obtain an approximate idea of the physical parameters.

3.2. MD Integrators

Construction of the HMC MD integrator was straightforward. On the larger lattices, we used three additional pseudofermions, following the method of Hasenbusch [23] and multiple time-scale integration [33] with the gauge field time steps being eight times smaller than that for the the heaviest pseudofermion, which in turn had a time step eight times smaller than that of the lightest pseudofermion.

3.3. Linear Equation Solvers

For any choice of rational approximation and kernel operator there are several different approaches to the problem of inverting the chiral Dirac operator $\slashed{D}_\mu$ [10].

1. Introduce a five dimensional matrix that has $\slashed{D}_\mu$ as its Schur complement.
   (a) Use this to find the inverse of $\slashed{D}_\mu$ applied to a four dimensional pseudofermion field in order to compute the fermionic force for four dimensional MD.
   (b) Use the five dimensional matrix as part of a five dimensional MD scheme with five dimensional pseudofermions (this corresponds to the domain wall formalism).
2. Apply the inverse of $\slashed{D}_\mu$ to a four dimensional pseudofermion field by use of a nested solver.

In this paper we use a nested four dimensional solver, using a partial fraction representation with relaxation [34], GMRESR preconditioning [35], and deflation of about 70 kernel eigenvalues.

It is not in the remit of this paper to compare the performance of four- and five-dimensional solvers; our goal is to compare the effect of different rational approximations. We expect that the relative performance of such different

Volume	β	$10^4\xi$	N	a (fm)	am_π	k	Acc. %
$32^3 \times 64$	8.65	KLN	16	0.101(3)	0.203(3)	—	90
$32^3 \times 64$	8.7	1.93	16	0.094(4)	0.182(5)	35	100
$32^3 \times 64$	8.7	0.643	16	0.096(3)	0.182(1)	18	97
$32^3 \times 64$	8.7	0.193	16	0.095(3)	0.183(2)	9	92

Table 1: Parameter values for ensembles. β is the Lüscher-Weisz gauge coupling, a and m_π are the lattice spacing and pion mass respectively. We also show N, the degree of the rational approximation, and the HMC acceptance rate. For the Zolotarev Dirac operator we list ξ and k, the average number of eigenvalues of the kernel operator H that lie outside the optimal interval of the Zolotarev approximation. The first line corresponds to our KLN approximation ensembles.

approximations to the sign function should be similar with the use of five dimensional solvers such as used in the domain wall algorithm.

The parameters of the runs are given in Table 1. Pion masses were measured using the pseudo-scalar correlator, and the lattice spacing using r_0 [36, 37]. Because we only have a handful of configurations for each ensemble, it was difficult to get a reliable estimate of the statistical errors, particularly for the lattice spacing, and a larger study is required to get more accurate values. However, the physical pion masses, which are in the range $370 - 400$ MeV, and the lattice spacings agree across our ensembles to within the errors.

4. Results

4.1. HMC Acceptance Rate

The change in the value of the Hamiltonian ΔE for our HMC runs is shown in Figure 5. There are no large spikes, nor any large differences between the various different Dirac operators. The acceptance rate shown in Tables 1 are similar (albeit high) for each of the HMC runs. The ensembles used the same Hasenbusch masses, MD time steps and, for our performance tests, the same number of deflated eigenvalues. There is no indication of any MD instabilities.

4.2. Performance

The total time taken for the inversions required for each HMC run for the $32^3 \times 64$ lattices is shown in Table 2. The number of projected Wilson eigenvalues varied from trajectory to trajectory, with an average around 70. All the measurements were performed with identical time-steps, trajectory lengths, and an equal number of pseudofermion fields. While generating the configurations, the order of all the Zolotarev Dirac and KLN operators was held fixed at 16.

While a larger study is needed for definitive results, it is clear that the computational cost required for the Zolotarev and KLN runs does not differ by any substantial amount. The observed, small, difference is caused by a slower convergence for the inversion of the Zolotarev Dirac operators compared to the hyperbolic tangent operator, which is partially explained by the slightly heavier
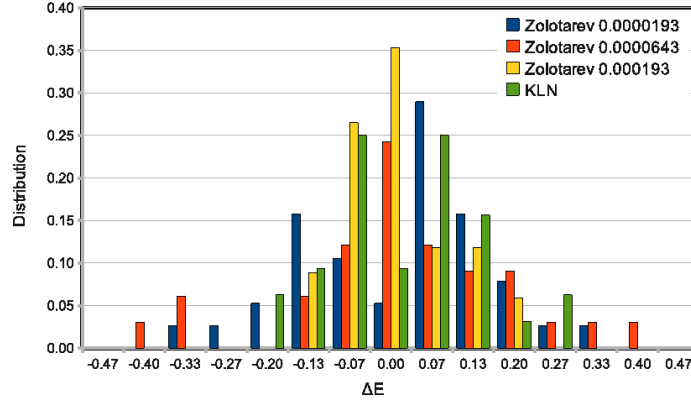
Figure 5: The distribution of the HMC energy difference for the KLN and three Zolotarev Dirac operators for the $32^3 \times 64$ lattice.

Volume	$10^4\xi$	Inverter	Eigenvalues
$32^3 \times 64$	KLN	$5,435$	$1,568$
$32^3 \times 64$	1.93	$6,291$	$1,627$
$32^3 \times 64$	0.643	$6,650$	$2,657$
$32^3 \times 64$	0.193	$6,194$	$1,641$

Table 2: Time (in seconds) spent in the two dominant parts of the HMC code in each trajectory.

13
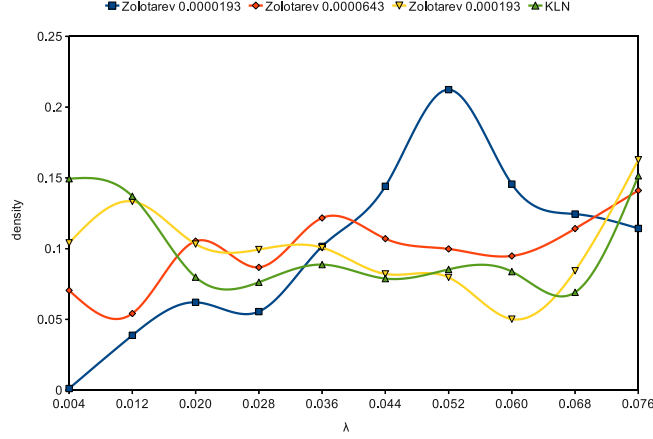
Figure 6: The distribution of the smallest eight eigenvalues of the kernel operator H during the molecular dynamics for the KLN and Zolotarev Dirac operators.

quark mass for the KLN. From these studies we conclude that the difference in cost between running Zolotarev and KLN fermions at equal degree and equal pion mass is at most about 20%. We must stress that these inversions were not performed at equal residual mass; to run at the same residual mass as the Zolotarev fermions, the cost for the KLN or domain wall fermions would be far greater (see §4.5).

4.3. Rate of topological charge changes

In Figure 6, we plot the distribution of the ten smallest eigenvalues of the kernel operator H as it evolved during the molecular dynamics. From this limited data we cannot reliably estimate the rate of topological charge change. The large fluctuations in the data are due to our small sample size; nonetheless the data is clear enough for a qualitative picture to emerge. Our data is consistent with our expectation that for the Zolotarev with largest ξ there is no suppression of the small kernel eigenvalues compared with KLN; although there are indications that the eigenvalues of the Zolotarev with the smallest ξ might be suppressed. No difference can be seen from our data between the rate of topological tunnelling for the Zolotarev at large ξ and KLN Dirac operators.

The suppression of the small eigenvalues also aids the stability of the HMC algorithm and the locality of the Zolotarev operator.

4.4. Locality of Dirac operators

We checked the locality of the Dirac operator by measuring the exponential decay from a single source. All the Dirac operators are exponentially local, and there is no noticeable difference in locality between the overlap operator and the Dirac operators studied here.
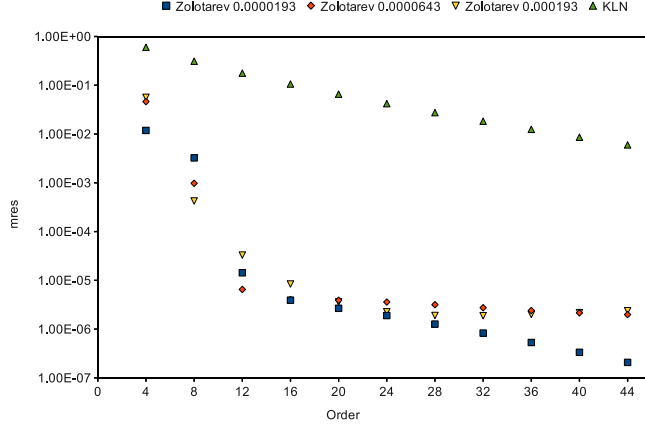
14

Figure 7: The residual mass m_{res} plotted as a function of the degree N of the KLN and three Zolotarev approximations, as measured on the $32^3 \times 64$ lattices.

4.5. Chiral symmetry breaking

The residual mass m_{res}, as defined in §1.2, is shown for both the KLN and all three Zolotarev Dirac operators as a function of the degree of the rational approximation in Figure 7. In this plot, m_{res} has been averaged over all configurations. It is immediately seen that both Zolotarev Dirac operators exhibit significantly smaller residual masses than the KLN operator for all but the lowest degrees. At the order used in our tests, $N = 16$, equivalent to $L_S = 16$ for domain wall fermions, the improvement is four orders of magnitude. We also observe a striking difference in the way in which the chiral symmetry is broken for the KLN and the Zolotarev Dirac operators. For the KLN Dirac operator the residual mass decreases steadily. For the Zolotarev Dirac operator it decreases rapidly up to around degree 16, and, for the larger Zolotarev ranges reaches a plateau, where the fluctuations are dominated by statistical noise, and for the smallest range continues to decrease but at a slower rate as the order is increased. Moreover, while m_{res} is roughly constant between configurations for the KLN Dirac operator, there is a large fluctuation between configurations for the Zolotarev Dirac operator. There is therefore little point in running a Zolotarev approximation, at least for the parameter values considered here, with degrees $N > 16$.

The pattern of chiral symmetry breaking for the Zolotarev Dirac operator can be easily explained: there are two contributions to the violation of the Ginsparg–Wilson relation for the Zolotarev Dirac operator, from the imperfection of the approximation to the matrix sign function within the interval on which the Zolotarev approximation is optimal (the "bulk"), and the contribution from the small eigenvalues below $\xi \|H\|$. The first contribution decreases rapidly with N, while the second contribution decreases considerably more slowly, if at
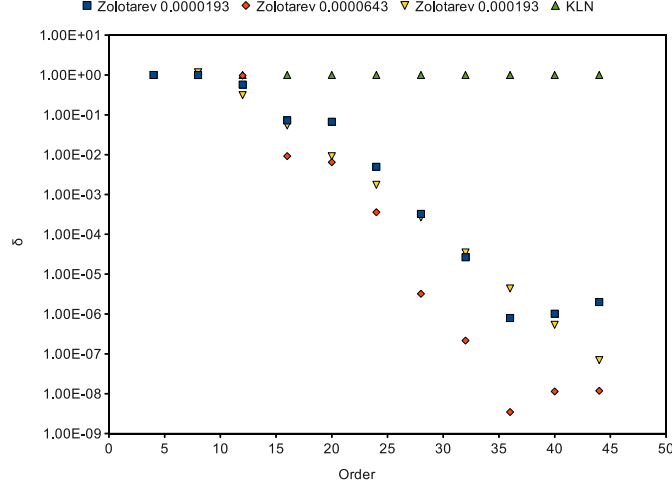
15

Figure 8: The ratio of the contributions to m_{res} from bulk modes to that from low modes as a function of the degree of the rational function approximation N for the KLN and three Zolotarev Dirac operators, measured on the $32^3 \times 64$ lattices.

all. For small N m_{res} is dominated by the bulk, while for large N it is dominated by the low modes.

These two sources of explicit chiral symmetry violation are illustrated in Figure 8, in which we have separated the contributions to m_{res} from the bulk modes from the modes below $\xi \|H\|$ and plotted their ratio, defined as

$$\delta = \frac{\sum_{x,x'} \left[\slashed{D}_\mu^{\dagger}{}^{-1} \gamma_5 \Delta \left(1 - \sum_i |\psi_i\rangle\langle\psi_i|\right) \slashed{D}_\mu^{-1} \right]_{x,x'}}{\sum_{x,x'} \left[\slashed{D}_\mu^{\dagger}{}^{-1} \gamma_5 \Delta \left(\sum_i |\psi_i\rangle\langle\psi_i|\right) \slashed{D}_\mu^{-1} \right]_{x,x'}}, \tag{7}$$

for for the lowest orthonormal eigenvectors, ψ_i, of the kernel operator H. It can be seen that for the Zolotarev approximation for order $N \geq 16$ the chiral symmetry breaking is almost entirely caused by the lowest eigenvalues of the kernel operator, while for lower N the bulk eigenvalues have a larger contribution. It is thus clear that the plateau in m_{res} for the $\xi\|H\| = 0.0000643$ ensemble is caused by the low modes, whereas the $\xi\|H\| = 0.0000193$ ensemble has no such small eigenvalues, and correspondingly the residual mass gradually improves as the order of the rational approximation improves and the approximation for the matrix sign function gets better for the bulk modes.

5. Conclusions

Our main conclusion is that the Zolotarev Dirac operator seems to provide a significant improvement over both domain wall and overlap fermions for computations with light fermions on fine lattices.

16

On coarse lattices contributions from the small eigenvalues subspace of the kernel operator dominate the residual mass. Some of these eigenvectors are topological, whereas others are lattice artefacts ("defects"). Various mechanisms have been suggested to suppress such defects without affecting the underlying continuum physics by increasing autocorrelations for topology change. To the extent that such methods are effective the contributions to the residual mass from the bulk will still be significantly reduced by the Zolotarev Dirac operator.

5.1. Comparison with Domain Wall fermions

The principal advantage that the Zolotarev Dirac operator has over domain wall fermions is that it gives a significantly smaller residual mass for the same computational effort. In particular, this allows for simulations at smaller lattice spacing than are currently possible with domain wall fermions. The second advantage is that the Zolotarev Dirac operator can be tuned to balance the performance of the HMC, the tunnelling rate and the residual mass.

There are still questions which need to be addressed in future work. In particular, we have not addressed the question of whether using five dimensional (as used by Chiu and collaborators [16]) or four dimensional pseudofermions as in this work is superior. Should the five dimensional inversion prove superior to the nested four dimensional inversion (the comparison in [38] used a considerably sub-optimal nested 4D algorithm, so this question remains open), then this can easily incorporated into the four dimensional algorithm, so it seems unlikely that there is much difference between the two. However, until a direct comparison is made in a future work, no definite statement can be made in favour of either formulation.

5.2. Comparison with Overlap fermions

The Zolotarev Dirac operator has the advantage over overlap fermions that it is faster. An exact overlap calculation requires an accurate resolution of the matrix sign function during the molecular dynamics, which in practice requires the transmission/reflection algorithm [25, 26, 20]. To allow frequent topological charge changes, the overlap HMC algorithm has to be further refined and carefully tuned [22], leading to approximately a doubling of the cost per trajectory. Furthermore, while these refinements allow topological charge changes every few trajectories, there may still be longer autocorrelations than with the Zolotarev Dirac operator. This cost can be removed by adding unphysical terms to the action [39], thus forbidding both topological tunnelling and kernel eigenvalues close to zero. The effect of this with regards to possible artefacts and ergodicity is, however, unclear.

Acknowledgements

References

[1] H. B. Nielsen, M. Ninomiya, Absence of neutrinos on a lattice. 1. Proof by homotopy theory, Nucl. Phys B185 (1981) 20. 1

[2] M. Lüscher, Exact chiral symmetry on the lattice and the Ginsparg–Wilson relation, Phys. Lett. B428 (1998) 342–345. `arXiv:hep-lat/9802011`. 1, 1.1

[3] P. H. Ginsparg, K. G. Wilson, A remnant of chiral symetry on the lattice, Phys. Rev. D25 (1982) 2649. 1, 1.1

[4] R. Narayanan, H. Neuberger, Chiral determinant as an overlap of two vacua, Nucl. Phys. B412 (1994) 574–606. `arXiv:hep-lat/9307006`. 1

[5] H. Neuberger, A practical implementation of the overlap-dirac operator, Phys. Rev. Lett. 81 (1998) 4060–4062. `arXiv:hep-lat/9806025`. 1, 1.4, 2.2

[6] D. B. Kaplan, A method for simulating chiral fermions on the lattice, Phys. Lett. B288 (1992) 342–347. `arXiv:hep-lat/9206013`. 1

[7] C. Kenney, A. Laub, A hyperbolic tangent identity and the geometry of Padé sign function iterations, Numer. Algorithms 7 (1994) 111–128. 1, 1.4, 2.2

[8] H. Neuberger, Exactly massless quarks on the lattice, Phys. Lett. B417 (1998) 141–144. `arXiv:hep-lat/9707022`. 1.1

[9] H. Neuberger, More about exactly massless quarks on the lattice, Phys. Lett. B427 (1998) 353–355. `arXiv:hep-lat/9801031`, `doi:10.1016/S0370-2693(98)00355-4`. 1.1

[10] A. D. Kennedy, Algorithms for dynamical fermions, in: Y. Kuramashi (Ed.), Perspectives in Lattice QCD, World Scientific, 2006, pp. 15–82. `arXiv:hep-lat/0607038`. 1.1, 2.2, 3.3

[11] E. I. Zolotarev, Application of elliptic functions to questions of functions deviating least and most from zero, Zap. Imp. Akad. Nauk St. Petersburg 30 (1877) 5. 2.1

[12] J. van den Eshof, A. Frommer, T. Lippert, K. Schilling, H. A. van der Vorst, Numerical methods for the QCD overlap operator. I: Sign function and error bounds, Comput. Phys. Commun. 146 (2002) 203–224. `arXiv:hep-lat/0202025`. 2.1

[13] A. D. Kennedy, Approximation theory for matrices, in: A. C. Kalloniatis, D. B. Leinweber, A. Williams (Eds.), LHP 2003, Vol. 128C of Nuclear Physics B (Proceedings Supplements), 2004, pp. 107–116. `arXiv:hep-lat/0402037`. 2.1

[14] A. D. Kennedy, Fast evaluation of Zolotarev coefficients, in: Boriçi et al. [41], pp. 169–190. `arXiv:hep-lat/0402038`. 2.1

[15] A. Boriçi, Computational methods for the fermion determinant and the link between overlap and domain wall fermions, in: Boriçi et al. [41], pp. 23–59. `arXiv:hep-lat/0402035`. 2.2

[16] T.-W. Chiu, Optimal domain-wall fermions, Phys. Rev. Lett. 90 (2003) 071601. `arXiv:hep-lat/0209153`. 2.2, 5.1

[17] S. Duane, A. D. Kennedy, B. Pendleton, D. Roweth, Hybrid Monte Carlo, Phys. Lett. B195 (1987) 216. 2.3

[18] N. Cundy, Small Wilson Dirac operator eigenvector mixing in dynamical overlap Hybrid Monte Carlo, Comput. Phys. Commun. 180 (2009) 180–191. `arXiv:0706.1971, doi:10.1016/j.cpc.2008.09.007`. 2.3, 2.3

[19] A. Bode, U. M. Heller, R. G. Edwards, R. Narayanan, First experiences with HMC for dynamical overlap fermions, in: Dubna 1999, Lattice fermions and structure of the vacuum, 1999, pp. 65–68, poster presented at NATO Advanced Research Workshop on Lattice Fermions and Structure of the Vacuum, Dubna, Russia, 5-9 Oct 1999. `arXiv:hep-lat/9912043`. 2.3

[20] N. Cundy, Current status of dynamical overlap project, Nucl. Phys. Proc. Suppl. 153 (2006) 54–61. `arXiv:hep-lat/0511047`. 2.3, 2.3, 5.2

[21] G. I. Egri, Z. Fodor, S. D. Katz, K. K. Szabo, Topology with dynamical overlap fermions, JHEP 01 (2006) 049. `arXiv:hep-lat/0510117`. 3

[22] N. Cundy, S. Krieg, T. Lippert, A. Schäfer, Topological tunneling with dynamical overlap fermions, Comput. Phys. Commun. 180 (2009) 201–208. `arXiv:0803.0294, doi:10.1016/j.cpc.2008.09.010`. 3, 2.3, 5.2

[23] M. Hasenbusch, Speeding up the Hybrid Monte Carlo algorithm for dynamical fermions, Phys. Lett. B519 (2001) 177–182. `arXiv:hep-lat/0107019`. 2.3, 3.2

[24] M. A. Clark, A. D. Kennedy, Accelerating dynamical fermion computations using the Rational Hybrid Monte Carlo (RHMC) algorithm with multiple pseudofermion fields, Phys. Rev. Lett. 98 (2007) 051601–051604. `arXiv:hep-lat/0608015`. 2.3

[25] Z. Fodor, S. D. Katz, K. K. Szabo, Dynamical overlap fermions, results with Hybrid Monte Carlo algorithm, JHEP 08 (2004) 003. `arXiv:hep-lat/0311010`. 2.3, 5.2

[26] N. Cundy, et al., Numerical methods for the QCD overlap operator. IV: Hybrid Monte Carlo, Comput. Phys. Commun. 180 (2009) 26–54. `arXiv:hep-lat/0502007`, `doi:10.1016/j.cpc.2008.08.006`. 2.3, 5.2

[27] P. J. Moran, D. B. Leinweber, Over-improved stout-link smearing, Phys. Rev. D77 (2008) 094501. `arXiv:0801.1165`, `doi:10.1103/PhysRevD.77.094501`. 3.1

[28] C. Morningstar, M. J. Peardon, Analytic smearing of $SU(3)$ link variables in lattice QCD, Phys. Rev. D69 (2004) 054501. `arXiv:hep-lat/0311018`. 3.1

[29] M. Lüscher, P. Weisz, On-shell improved lattice gauge theories, Commun Math Phys 97 (1985) 59. `doi:10.1007/BF01206178`. 3.1

[30] G. Curci, P. Menotti, G. Paffuti, Symanzik's improved Lagrangian for lattice gauge theory, Phys. Lett. B130 (1983) 205. `doi:10.1016/0370-2693(83)91043-2`. 3.1

[31] M. Lüscher, P. Weisz, Computation of the action for on-shell improved lattice gauge theories at weak coupling, Phys. Lett. B158 (1985) 250. `doi:10.1016/0370-2693(85)90966-9`. 3.1

[32] J. R. Snippe, Computation of the one-loop Symanzik coefficients for the square action, Nucl. Phys. B498 (1997) 347–396. `arXiv:hep-lat/9701002`, `doi:10.1016/S0550-3213(97)00270-8`. 3.1

[33] J. C. Sexton, D. H. Weingarten, Hamiltonian evolution for the Hybrid Monte Carlo algorithm, Nucl. Phys. B380 (1992) 665–678. `doi:10.1016/0550-3213(92)90263-B`. 3.2

[34] G. Arnold, et al., Numerical methods for the QCD overlap operator. II: Optimal Krylov subspace methods, Lecture Notes in Computational Science and Engineering 47 (2005) 153. `arXiv:hep-lat/0311025`. 3.3

[35] N. Cundy, et al., Numerical methods for the QCD overlap operator. III: Nested iterations, Comput. Phys. Commun. 165 (2005) 221–242. `arXiv:hep-lat/0405003`. 3.3

[36] R. Sommer, A new way to set the energy scale in lattice gauge theories and its applications to the static force and $alpha_s$ in $su(2)$ Yang–Mills theory, Nucl. Phys. B411 (1994) 839–854. `arXiv:hep-lat/9310022`. 3.3

[37] C. R. Allton, et al., Effects of non-perturbatively improved dynamical fermions in QCD at fixed lattice spacing, Phys. Rev. D65 (2002) 054502. `arXiv:hep-lat/0107021`, `doi:10.1103/PhysRevD.65.054502`. 3.3

[38] S. Hashimoto, et al., Lattice simulation of 2+1 flavors of overlap light quarks, PoS LAT2007 (2007) 101. `arXiv:0710.2730`. 5.1

[39] S. Hashimoto, et al., Dynamical overlap fermion at fixed topology, PoS LAT2006 (2006) 052. `arXiv:hep-lat/0610011`. 5.2

[40] D. J. Antonio, et al., First results from 2+1-flavor domain wall QCD: Mass spectrum, topology change and chiral symmetry with L(s) = 8, Phys. Rev. D75 (2007) 114501. `arXiv:hep-lat/0612005`, `doi:10.1103/PhysRevD.75.114501`. 5.2

[41] A. Boriçi, B. J. Andreas Frommer, A. D. Kennedy, B. J. Pendleton (Eds.), QCD and Numerical Analysis III, Springer, 2005. 5.2